Car Accidents Severity in Seattle

Nastaran Dashti

October 22, 2020

1 Introduction

Road traffic injuries are within the top ten cause of death worldwide[1]. Every year, over 1.35 million people lose their life on roads which 150 thousands of them are from the United States. Over half of all road traffic deaths are among pedestrians, cyclists and motorcyclists.

Seattle, one of the fast-growing city in the West Coast of the US with a population of 3.95 million [2], recorded the highest number of car accidents in the entire country [3] in 2015. To reduce the number and severity of accidents in the future, we need to study the car accident data over previous years. Here, an analysis and prediction of the severity and number of accidents using various factors such as weather, road condition, location and sobriety of driver, and many other parameters that I will elaborate later in data section, is reported.

This report will guide first, the Seattle government on how to manage and reduce severity of car accidents, and second, the drivers on how to prevent getting into a car accident by changing either their travel plan or driving behavior.

2 Data

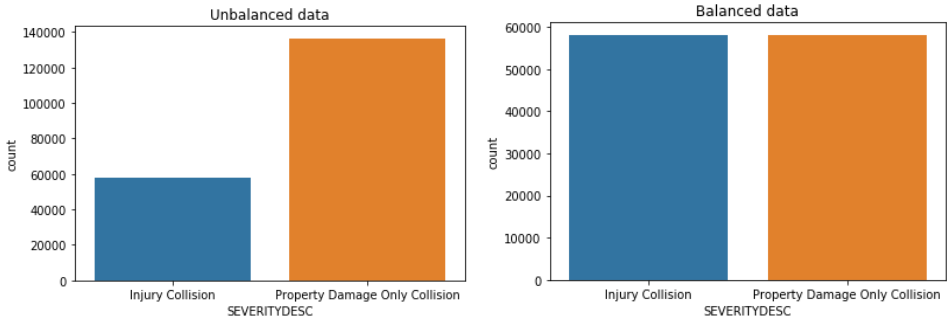
The data is provided by SDOT Traffic Management Division [4] from 2014 to present. Here, for each accident around 37 features are considered. Some of them are listed below:

- LOCATION: Latitude (X) and longitude (Y) of accident.
- ADDRTYPE: Collision address type included alley, block, intersection.
- JUNCTIONTYPE: Category of junction at which collision took place.

- COLLISIONTYPE: Collision type such as parked Car, Angles and etc.
- WEATHER: A description of the weather conditions during the time of the collision.
- ROADCOND: The condition of the road during the collision.
- LIGHTCOND: The light conditions during the collision.
- UNDERINFL: Driver involved was under the influence of drugs or alcohol.
- PERSONCOUNT: The total number of people involved in the collision.
- INCDTTM: The date and time of the incident.

To build a good model, the dataset should be rich and contains many observations. Unfortunately, EXCEPTRSNCODE, EXCEPTRSNDESC, PEDROWNOTGRNT, SPEEDING, INATTENTIONIND and INTKEY have a high number of missing data, therefore I decided not to include them in the analysis.

The features SEVERITYCODE/SEVERITYDESC, representing different levels of severity caused by the accident, are used as a target. As you can see in the Fig 1(a), the majority of the accidents are related to the property damage and there is no data on the fatality. This can be interpreted in two ways; there is either a lack of information on fatality or no serious accident occurred during those years.



(a) (b)

Figure 1: The severity of the collision with (a) unbalanced, and (b) balanced data.

In order to improve the accuracy of the predictive machine learning models, the data needs to be balanced between the two categories. For this purpose, the resample library has been used to reduce the number of property damage, as it is shown in Fig 1(b).

Finally, a supervised learning model will be used to come up with a formula that can predict the severity of an accident based on the inputs.

References

[1] What do people die from? (2020). URL `https://ourworldindata.org/what-does-the-world-die-from`.

[2] Seattle - Wikipedia (2020). URL `https://en.wikipedia.org/w/index.php?title=Seattle&oldid=984462802`.

[3] Washington State Car Accident Statistics, journal = Colburn Law, year = 2020, month = Mar, url = https://www.colburnlaw.com/seattle-traffic-accidents.

[4] Traffic Management Division of Seattle, WA (2020). URL `https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv`.